

# ERROR TREATMENT IN THE EFL WRITING CLASS: RED PEN METHOD VERSUS REMEDIAL INSTRUCTION

By

Dr. MOHAMMAD ALI SALMANI-NODOUSHAN\*

## ABSTRACT

*In a study conducted to see which method of error treatment was more effective in EFL writing classes, 288 Iranian EFL learners took the TOEFL test to be grouped in two homogeneous classes. Each student in each group wrote a paragraph on a general topic which was proofread for mistakes/errors by three experienced EFL writing teachers (i.e., Pretest). One group received Red Pen treatment (RPM) and the other Remedial Instruction treatment (RIM). After a two-week interval, both groups repeated the same writing assignment, proofread by the same teachers (Post-test). A Mixed Between-Within Subjects Analysis of Variance (SPANOVA) was conducted to analyze the effect of two different types of treatment (i.e., RPM, and RIM). Results, after analysis of the data, indicated that the main effect was significant for time but not for group. It was further noticed that the interaction effect was also significant. The RPM method, although not statistically significant, was slightly more effective in enhancing EFL written performance than the RIM method.*

**Keywords:** *EFL Writing; Classroom Interaction; Error Treatment; Red Pen Method; Remedial Instruction*

## INTRODUCTION

EFL writing teachers employ different methods of error treatment in their classes. Two of the most popular methods of error treatment in Iran are (a) The Red Pen Method (RPM), and (b) the Remedial Instruction Method (RIM). In RPM, the teacher writes notes in red ink through which s/he draws students' attention to their errors, and asks for a revision. In RIM, on the other hand, s/he focuses on students' errors in follow-up class hours in which the teacher teaches the students how to avoid those errors in later writing assignments. This study aimed at finding the answer to the question: "Which of these two methods of error treatment is more effective?"

### 1. Background

Second language writing has been the subject matter of many researches until now. There have been different points of view about types of error treatment in writing English as a second language. Some of these have approved the use of Red Pen Method (RPM) as a type of error treatment. This method is said to attract students' attention, and once they notice their mistakes, they try to avoid those specific mistakes in later assignments. However, the red pen method, as one method of treatment, has been criticized, too. As Semke (1984) puts it, correction does not increase writing accuracy, writing

fluency, or general language proficiency, and may have a negative effect on students' attitudes.

Along the same lines, in agreement with Cane and Cane (1990), Porte (1993) states that many methods of giving feedback on writing have tended to concentrate on teacher-initiated correction with the inevitable display of "codes" (usually in red ink) that aim to point students in the direction of their error or mistake by commenting in margins; Gwin (1991) noticed that this kind of commenting was sometimes implemented by color-coding. Martin, et al. (1976) noticed that the results of such systems were often less than satisfactory, with the teacher spending more time dealing with the surface features of spelling, punctuation and handwriting than other things which were either useful or desirable.

Approaching marking in such a way, coupled with EFL/ESL teachers' customary professionalism, many would find themselves wishing to spend more and more time with each student's problems, providing more exhaustive feedback. As Gwin (1991) points out, such teachers usually insist that compositions be double spaced, to give them plenty of room to write their comments. While such attention to detail is laudable, teachers cannot help but feel that they simply do not have the time to devote to such painstaking marking, since they are already

overloaded with their works.

On the other hand, using different types of treatment (e.g., Remedial instruction Method, Red Pen Method, etc.) to increase writing performance has its own advocates. Fox (1979) explains about a sixteen-week study to investigate the effects that two methods of teaching - writing had on writing apprehension and on overall quality and length of student writing involved over one hundred college freshmen enrolled in English Composition classes. Except for the methods of writing instruction, all other conditions for the experimental and control groups, such as class hours, number of words assigned and choice of topics were the same. By administering Daly and Miller's Writing Apprehension Test before and after the treatment, Fox (1979) observed that both groups had reduced in writing apprehension, but the experimental group, which had had the experimental treatment, showed a better result.

Several studies have been conducted on the effect of treatment on EFL students' writing performance. They all show that there is a rather significant change in writing ability after a certain type of treatment is performed. One issue about treatment is whether all types of treatment have the same effect, or some are more effective. This aims to check the effect of two different types of treatment on EFL students' writing performance: (a) the Red Pen Method (RPM), and (b) the Remedial instruction Method (RIM).

## 2. Methodology

### 2.1. Materials

The general proficiency test of English used in this study consisted of two booklets. Both booklets were selected questions from the book entitled *Longman Complete Course for the TOEFL Test* (Phillips, 2001). The first booklet contained grammar, and vocabulary questions. The second booklet contained reading comprehension passages and questions.

- The Grammar and Vocabulary Booklet: It contained forty grammar, and forty vocabulary questions to assess students' knowledge of English grammar and vocabulary.

- The Reading Comprehension Booklet: The second booklet with five standard passages, each followed by ten relevant comprehension questions, was used to test reading comprehension ability of the students.

*Advanced Writing*: this is the title of a practical course book on academic paragraph writing; it was introduced to all the participants in the study.

### 2.2. Participants and procedures

Based on TOEFL scores, a group of female Iranian students ( $N=288$ ) were chosen from among an original population of 362 EFL students for the study. They were mostly teenagers and young adults, ranging in age from fifteen to thirty years (mean age=23.2). They were all EFL students, means that they had learnt English in educational settings not in naturalistic settings or in the environment.

It was important to know their language proficiency before they could enter the study. To this end, a general proficiency test (i.e., the TOEFL Test) was administered. To reduce the effect of all the intervening variables, they were not given prior information about the exact date of test administration; so that could not prepare for it. After administering the TOEFL test, they were grouped into four different levels of language proficiency. These groups consisted of advanced, upper intermediate, lower intermediate, and beginning students. Table 1 displays the descriptive statistics for four proficiency groups.

The author had a notion that sex could be a possible intervening variable. In addition, most of the students in the population were female. Therefore, it was decided to draw the sample from among female students. As such, all the participants were female so that the variable of sex could be controlled.

	Frequency	Percent	Valid Percent	Cumulative Percent
Beginner	80	27.8	27.8	27.8
Lower Intermediate	80	27.8	27.8	55.6
Upper Intermediate	48	16.7	16.7	72.2
Advanced	80	27.8	27.8	100.0
Total	288	100.0	100.0	

Table 1. Distribution of Participants across Proficiency Groups

Type of Method	N	Mean	Std. Deviation	t	Sig.
Red Pen Method	144	44.4444	16.11921	-1.469	0.143
Remedial Instruction Method	144	47.0556	13.95887		

**Table 2. Descriptive and T-Test Statistics for RPM and RIM Groups**

It was now important to assign subjects to the two treatment groups in such a way as to ensure inter-group homogeneity in terms of language proficiency. To this end, a matching technique was used. This was done to ensure maximum correspondence between the RPM ( $n=144$ ) and RIM ( $n=144$ ) groups in terms of participants' level of language proficiency; for each TOEFL score in the RPM group, a corresponding score in the RIM group was desired, so that there was a one-to-one correspondence between TOEFL scores in RPM and RIM groups; that is, each TOEFL score in the RPM group had a counterpart in the RIM group. This was consolidated by running an independent samples t-test for the RPM and RIM groups with language proficiency (i.e., TOEFL scores) as the dependent variable ( $\alpha=0.05$ ) as in table 2.

Both groups were then asked to write a paragraph on the general topic "Why did you choose to study English?", as the pretest to assess their writing performance before either of the treatment methods (i.e., RPM or RIM) could be used. The paragraphs were proofread by three EFL writing instructors (average years of teaching experience=5.3) and the errors were identified. They also assigned scores to each student. As such, each student had three scores, the average of which was considered the pretest score for her.

The errors were indicated to the RPM group by comments in red ink in the margins, and each RPM participant was asked to revise her paragraph in which she tried to correct those errors indicated to her in red ink; the period for revision was a due of maxim of two weeks time. The revised paragraphs were retained as the post-test corpus for the RPM group.

As for the RIM group, the participants were asked to participate in Remedial Instruction Classes in which they

were taught the correct forms of the errors they had made in their paragraphs. They did not see their paragraphs and the instructor did not tell each individual what errors she had made. Rather, the RIM group, as a whole, was given remedial instruction. After six two-hour class sessions held over a two-week time, the RIM participants were asked to rewrite their paragraphs on the same topic. The rewritten paragraphs were retained as the post-test corpus for the RIM group.

The same EFL instructors proofread the post-test paragraphs and assigned scores to them. Once more, each individual received three scores, and the average of which was retained as her post-test score.

### 3. Results

A Mixed Between-Within Subjects Analysis of Variance (SPANOVA) (See Pallant, 2001) was conducted to analyze the effect of two different types of treatment (i.e., RPM, and RIM) on the writing performance of EFL students. This was done to see if there were main effects for each of the independent variables (i.e., main effect for subject groups and main effect for time), and also for their interaction to tell if the change in writing performance over time was different for the two groups.

It was necessary to check for *Homogeneity of intercorrelations* to see if for each of the levels of the between-subjects variable (i.e., type of treatment) and the pattern of intercorrelations among the levels of within-subjects variable (i.e., time) were the same. To test this assumption, Box's M statistic with the more conservative  $\alpha$  level of .001 was used with the hope that the statistic would not be significant (i.e., that the  $p$  level would be greater than 0.001). In other words, Box's M statistic tests the null hypothesis that the observed covariance

Box's M	10.543
F	3.488
df1	3
df2	14723280.000
Sig.	.015
Design:	
Intercept+Treatment	
Within Subjects Design:	
Time	

**Table 3. Box's Test of Equality of Covariance Matrices**

Effect		Value	F	Sig.	Partial Eta <sup>2</sup>
Time	Pillai's Trace	.272	107.091(b)	.000	.272
	Wilks' Lambda	.728	107.091(b)	.000	.272
	Hotelling's Trace	.374	107.091(b)	.000	.272
	Roy's Largest Root	.374	107.091(b)	.000	.272
Time * Treatment	Pillai's Trace	.166	57.122(b)	.000	.166
	Wilks' Lambda	.834	57.122(b)	.000	.166
	Hotelling's Trace	.200	57.122(b)	.000	.166
	Roy's Largest Root	.200	57.122(b)	.000	.166

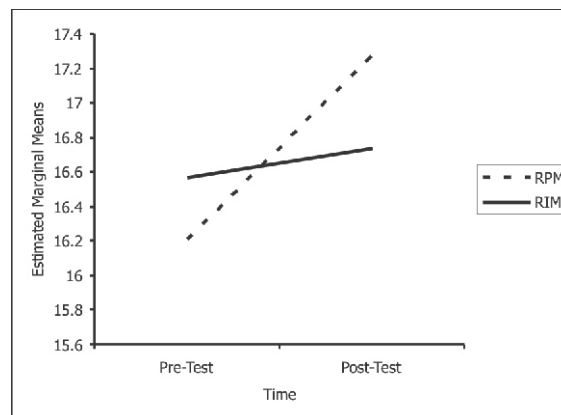
Computed using alpha = .01  
(Exact statistic, Design: Intercept  
+ Treatment, Within Subjects Design: Time)

**Table 4. Multivariate Tests**

matrices of the dependent variables are equal across groups. Table 3 displays the result and indicates that this assumption was met ( $Sig. = 0.015$ ).

A look at the Multivariate Tests table also indicated that there was a change in writing performance across time. The main effect for time was significant. There was also an indication that the two groups were also different in terms of writing performance across time. The main effect for the interaction between time and type of treatment was also significant. These findings are indicated by Wilks' Lambda values and the associated probability values given in the column labeled as  $Sig.$  in Table 4.

Based on the values in the Wilks' Lambda's part of the "Multivariate Tests" (Table 4) it was found that there was a statistically significant change in writing performance as a result of treatment. The value for Wilks' Lambda for time was 0.728, with a  $Sig.$  value of .000 (which means  $p < .0001$ ). Because the  $p$  value was less than .01, it was concluded that there was a statistically significant effect for time. This suggested that there was a change in writing performance across time; technically speaking, it showed the effect of treatment on writing ability. The value for partial Eta squared for time was 0.272. Using the commonly used guidelines proposed by Cohen's (1988) (0.01=small effect, 0.06=moderate effect, and 0.14=large effect), this result suggested a very large effect size for time.



**Figure 1. Comparison of gains in mean performance across subject groups.**

Furthermore, the value for Wilks' Lambda for time-treatment interaction was 0.834, with a  $Sig.$  value of .000 (which means  $p < .0001$ ). Because the  $p$  value was less than .01, it was concluded that there was a statistically significant effect for time-treatment interaction. The partial Eta squared value for the interaction effect was 0.166. This suggests a very large effect for time-treatment interaction. This means that there was not the same change in writing performance over time for the two treatment groups. In other words, gain in writing performance for the RPM group was not statistically the same as that for the RIM group. Figure 1 visualizes this difference in gains in writing performance across subject groups.

As Figure 1 indicates, the RPM group showed a greater gain in writing performance than the RIM group. Table 5 presents the descriptive statistics for the two treatment groups across time.

As Table 5 indicates, the pre-test mean for RPM was 16.20 while the post test mean was 17.27; the pre-test mean for

	Type of Treatment	Mean	Std. Deviation	N
Pre-test Score	RPM	16.2083	1.30156	144
	RIM	16.5694	1.08649	144
Post-test Score	RPM	17.2778	1.22300	144
	RIM	16.7361	1.23603	144

**Table 5. Descriptive Statistics for Treatment Groups across Time**

Source	Type II Sum of Squares	df	Mean Square	F	Sig.	Partial Eta <sup>2</sup>
Intercept	160600.563	1	160600.563	65945.219	.000	.996
Treatment	1.174	1	1.174	.482	.488	.002
Error	696.514	286	2.435			

Transformed Variable: Average  
Computed using alpha = .01

**Table 6. Tests of Between-Subjects Effects**

RIM was 16.56 whereas the post test mean was 16.73. The mean change was mathematically small but the statistical significance was checked from the data as displayed in Table 6.

As Table 6 indicates, the Sig. value for treatment was not statistically significant (Sig. = 0.488). The Sig. value was not less than the alpha level of 0.01, therefore it was concluded that the main effect for group was not significant. That is, there was no significant difference in gains in writing performance for the two groups (those who received RPM and those who received RIM). The effect size of the between-subject effect also supported this finding; the eta-squared value for treatment (or group) was 0.002. This is very minimal. It is therefore not surprising that it did not reach statistical significance.

## Discussion

This study was designed to see which intervention method (RPM or RIM) was more effective in enhancing EFL learners' writing performance across the two time periods (pre-intervention, and post-intervention). More specifically, it tried to see which treatment method was more effective.

The results after analysis of the data indicated that the main effect for time was significant. The positive mean difference for both groups as illustrated in Figure 1 and Table 5 above, showed that both groups had gains in writing performance as a result of the treatment they had received; however, the comparison of the two methods of treatment (RIM versus RPM) indicated that there was no significant difference between the two. In other words, the main effect for group was not statistically significant. This means that both methods of treatment are almost equally effective in enhancing EFL writing performance; although RPM appears to be slightly better than RIM, this

was not large enough to reach statistical significance. Therefore, the EFL teacher may choose either of the treatment methods she/he prefers.

## Conclusion

This study tried to see which method of treatment (RPM or RIM) was more effective in EFL writing classes. It was found that both methods of error treatment were almost equally effective. However, the RPM method was slightly better than the RIM method. This might be due to the very fact that in RPM the student is placed on a journey in which she/he is expected to learn through a discovery procedure. In fact, this discovery procedure may result in a deeper learning compared to what happens in the more or less deductive RIM method, in which the teacher decides what to tell the students and what not to.

Maybe this could have resulted from another source. In RPM, errors made by each individual were indicated to her in red ink; in RIM, on the other hand, errors were indicated to the class as a whole regardless of whether an individual had made them. As such, RPM is rather individualized whereas RIM is not. This individualized nature of RPM is also a probable cause for greater gain by the RPM group compared to the RIM group.

## Further Study

It should also be noted that all the participants in this study were female. This may indicate that the same results might not be gained from male samples. It is a good idea to replicate the same study with taking male students as samples to see if the same patterns appear.

## References

- [1]. Birjandi, P., Alavi, S. M., & Salmani-Nodoushan, M. A. (2004). *Advanced writing*. Tehran: Zabankadeh Publications.
- [2]. Cane, P., & Cane, E. (1990). *Write for first certificate*. London: Nelson Books.
- [3]. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- [4]. Fox, R. F. (1979). Treatment of writing apprehension and its effects on composition. Research paper (website; source N/A).

- [5]. Gwin, T. (1991). Giving students the write idea: A way to provide feedback on writing. *English Teaching Forum*, 29 (3), 2-5.
- [6]. Martin, N., D'Arcy, P., Newton, B., & Parker, R. (1976). *Writing and learning across the curriculum*, 11-16. Schools Council Writing across the Curriculum Project/University of London Institute of Education. Upper Montclair, N. J.: Boynton/Cook Publishers.
- [7]. Pallant, J. (2001). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows (Version 10)*. Buckingham: Open University Press.
- [8]. Phillips, D. (2001). *Longman complete course for the TOEFL test*. White Plains New York: Pearson Education
- [9]. Porte, G. K. (1993). Mistakes, Blanks, and Blank Checks. *Forum*, 29 (3).
- [10]. Semke, H. D. (1984). Effects of the Red Pen. *Foreign Language Annals*, 17 (3), pp. 195-202.
- [7]. Pallant, J. (2001). *SPSS survival manual: A step by step*

---

### ABOUT THE AUTHOR

\*English Department, College of Humanities, University of Zanjan, Iran

Dr. Mohammad Ali Salmani-Nodoushan is an assistant professor of TEFL at the English Department of University of Zanjan, Iran. His research interests include language testing in general, and testing English for Specific Purposes, Computer Adaptive Testing, and Performance Assessment in particular. He is currently a member of the editorial board of The Asian EFL Journal, The Linguistics Journal and i-manager's Journal of Educational Technology. Correspondence: [nodushan@ut.ac.ir](mailto:nodushan@ut.ac.ir) or [nodushan@znu.ac.ir](mailto:nodushan@znu.ac.ir)

